

Citation for published version:

Finus, M & Pintassilgo, P 2014, Formation of climate agreements: The role of uncertainty and learning. in *Toward a New Climate Agreement: Conflict, Resolution and Governance*. Taylor and Francis, pp. 29-43.

Publication date:
2014

Document Version
Peer reviewed version

[Link to publication](#)

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Formation of Climate Agreements: The Role of Uncertainty and Learning

MICHAEL FINUS AND PEDRO PINTASSILGO

Abstract

This chapter addresses the impact of uncertainty and learning for the success international climate agreements. The analysis is based on a simple a stylized coalition formation model, considering three learning scenarios and three types of uncertainty. We show, different from previous studies, that the “veil of uncertainty” is generally not conducive to the success of international agreements, even in a setting of strategic interaction. That is, we find that, usually “the more we learn the better it is”. Moreover, for those cases where more information leads to less successful agreements, we propose a mechanism to mitigate the negative impact of learning.

Keywords: climate change, self-enforcing agreements, transnational cooperation, uncertainty, learning

1. Introduction

Climate change is a major challenge to international cooperation, as emphasized for instance by the Stern Review and various IPCC reports (Stern 2006 and IPCC 2007) but also many others. One of the main problems of achieving cooperation under international climate agreements is free-riding. Countries have an incentive to adopt a non-cooperative behavior. Emission reduction constitutes a public good. No country can be excluded to benefit from the emission reduction of other countries. Moreover, by not contributing to emission reduction, a country saves on abatement cost. The literature on the formation of self-enforcing international environmental agreements (SEIEAs) studies the underlying incentive structure in detail, by considering various assumptions related to the behavior of countries and their objectives, the cost-benefit structure and many other economic features affecting the incentive structure of governments to join climate treaties, but also pointing to possibilities to mitigate the free-rider incentives. For surveys, see for instance Barrett (2003) and Finus (2003 and 2008).

Uncertainty is also an important element in climate change and hence determines the formation of international agreements. In fact, despite intensive research, there are still large uncertainties regarding the impact of greenhouse gases on the climate system and on caused environmental damages. In addition, predictions about abatement costs are difficult (IPCC 2007). These uncertainties may well have an impact on global climate change governance. For instance, the former US President George Bush used uncertainty (as an excuse?) as one argument for his decision to withdraw from the Kyoto Protocol. In a letter to Senators, dated March 13, 2001, as quoted by Kolstad (2007), he wrote: “I oppose the Kyoto Protocol ... we must be very careful not to take actions that could harm consumers. This is especially true given the incomplete state of scientific knowledge”.

A recent strand of literature has analyzed the formation of international agreements on climate change in the context of uncertainty (Kolstad 2007, Kolstad and Ulph 2008, 2011, Na and Shin 1998). The main conclusion is that the “veil of uncertainty” is conducive to the success of international climate agreements. That is, in a model that captures the strategic interaction among countries in climate change, more information through learning can lead to worse outcomes. Outcomes are measured in terms of

aggregate welfare (i.e. the sum of welfare levels in all countries) in the equilibria of a coalition formation game under various assumptions about the degree of learning. This result is certainly puzzling as it runs counter to the general wisdom that more/better information can never harm. This result is also somehow disturbing because it basically implies that all scientific research, aiming at reducing the uncertainty around climate change, would be counter-productive. This leads to three research questions which we would like to address in this paper.

1. What are the driving forces that generate the “negative result” about learning?
2. How general is this negative result?
3. Can the problem be fixed?

In answering these questions, it is helpful to point out that the papers cited above use stylized models and make a couple of simplifying assumptions to derive their results. Hence, one route to address our questions could be to set up a model with “more realism”. This route has been pursued in Dellink et al. (2008) and Dellink and Finus (2012) who use a calibrated climate change model with twelve world regions and determine stable coalitions based on a large set of Monte Carlo Simulations under various assumptions about the distribution of the key parameters. From their results, it appears that the negative impact of learning is less evident than in the purely theoretical papers. In fact, in most cases, full learning leads to better outcomes than partial or no learning. Another route is to modify and generalize previous theoretical models. This approach has been taken by Finus and Pintassilgo (2012 and 2013) of which we summarize the main findings in a non-technical way in this chapter. The advantage of the theoretical approach is that driving forces can be isolated and one can analyze the generality of the previous results in a systematic way. Nevertheless, we are well aware of a couple interesting extensions, which could add more realism to our model. These extensions are briefly discussed in the last section.

2. The Model

Coalition formation is usually modeled using game theory – a branch of mathematics that studies the strategic interaction between decision makers (“players”) by using various equilibrium concepts to predict the outcome of these interactions (Finus 2001). In the following, we first introduce the coalition formation game in section 2.1, then explain the three types uncertainty in section 2.2 and finally lay out the three scenarios of learning in section 2.3.

2.1 Coalition Formation Game

International environmental agreements are typically “single agreements”, meaning that countries decide either to join a treaty (in which case they are a member of the “coalition”) or to abstain (in which case they act as singletons). Moreover, participation is voluntary and membership is open to all, i.e. a country can neither be forced into nor excluded from participation. Therefore, we model coalition formation as a two-stage open membership single coalition game.¹ In the first stage, players (i.e. countries in our context) decide whether to join an agreement (i.e. a climate treaty in our context) or remain an outsider as a singleton. In the second stage, players choose their policy levels (i.e. abatement in our context). The game is solved backward, assuming that strategies in each stage must form a Nash equilibrium, i.e. they are mutual best replies.

This simple game has also been called cartel formation game with non-members called fringe players. It originates from the literature in industrial organization (d’Aspremont et al. 1983) and has been widely applied in this literature (see Bloch 2003 and Yi 1997 and 2003 for surveys) but also in the literature on self-enforcing international environmental agreements (see Barrett 2003 and Finus 2003 for surveys).

In the first stage, players’ membership decisions lead to a coalition structure, $K = \{S, I_{n-m}\}$, which is a partition of players, with n being the total number of players, m the size of coalition S , $m \leq n$, I_{n-m} denotes the $n-m$ singletons and N the set of players, $S \subseteq N$. Due to the simple structure of this coalition formation game, there can be at most one non-trivial coalition, with “non-trivial” referring to a coalition of

at least two players. Hence, we can simply talk about coalition S with the understanding that all players that are not in S are singletons, i.e. single players.ⁱⁱ Typically, we will denote a member of S by i and call it a signatory and a non-member of S by j and call it a non-signatory.

In the second stage, given that some coalition S has formed in the first stage, players choose their abatement levels q_i . The decision is based on the following payoff function:

$$(1) \quad \Pi_i = B_i \left(\sum_{k=1}^n q_k \right) - C_i(q_i), \quad i \in N$$

where $B_i(\bullet)$ is country i 's concave benefit function (i.e. benefits increase at a constant or decreasing rate) from global abatement (in the form of reduced damages, e.g. measured against some business-as-usual-scenario), with global abatement being the sum of all abatement and $C_i(\bullet)$ its convex abatement cost function (i.e. costs increase at a constant or increasing rate) from individual abatement. The global public good nature of abatement is captured by the benefit function which depends on the sum of all abatement contributions. For a start, we assume that all functions and their parameters are common knowledge and introduce uncertainty later on.

Working backward, we assume that the optimal economic strategies in the second stage are the Nash equilibrium of the game between coalition S , with its m members, and the $n-m$ singletons. The equilibrium is derived by assuming that coalition members maximize the aggregate payoff of their coalition whereas all singletons maximize their own payoff. That is, coalition members cooperate and the coalition internalize the externality among its members. In contrast, singletons behave selfishly, ignoring the externality they impose on others. The simultaneous solution of these maximization problems leads to the equilibrium abatement levels of signatories $q_i^*(S)$ and of non-signatories $q_j^*(S)$. The abatement levels depend on coalition S . If say a non-member k joins coalition S such that $S \cup \{k\}$, then the equilibrium abatement level of coalition members will increase and those of non-members will decrease or remain

constant, i.e. $q_i^*(S) < q_i^*(S \cup \{k\})$ and $q_j^*(S) \geq q_j^*(S \cup \{k\})$. It can be shown that total abatement will increase if a non-signatory k joins coalition S such that $S \cup \{k\}$ forms. Hence, non-members are better off under $S \cup \{k\}$ than under S as benefits from total abatement will be higher and their abatement costs will remain constant or will drop. That is, non-members benefit from more cooperation of others, which explains the strong free-rider incentive, which typically shows up in only small coalitions being stable.

If equilibrium abatement levels of signatories $q_i^*(S)$ and of non-signatories $q_j^*(S)$ are inserted in the payoff function (see (1) above), we derive equilibrium payoffs in the second stage of the coalition formation game, given that coalition S has formed, which are denoted by $\Pi_{i \in S}^*(S)$ and $\Pi_{j \notin S}^*(S)$, respectively.

In the first stage, stable coalitions are determined by invoking the stability concept of internal and external stability, which is de facto a Nash equilibrium in membership strategies.

(2) Internal stability: $\Pi_i^*(S) \geq \Pi_i^*(S \setminus \{i\}) \quad \forall i \in S$

(3) External stability: $\Pi_j^*(S) > \Pi_j^*(S \cup \{j\}) \quad \forall j \notin S$.

That is, no signatory should have an incentive to leave coalition S to become a non-signatory and no non-signatory should have an incentive to join coalition S . In order to avoid knife-edge cases, we assume that if players are indifferent between joining coalition S and remaining outside of S , they will join the coalition. Coalitions which are internally and externally stable are called stable. In case there is more than one stable coalition, we apply the Pareto-dominance selection criterion. That is, we delete those stable coalitions from our set of stable coalitions where at least one player could be made better off and no player worse off by moving to another stable coalition.

Up to now, we have assumed the absence of transfers. However, given the assumption of joint welfare maximization of coalition members and the fact that we allow for asymmetric payoff functions, it is perceivable that coalition members share their total payoff $\Pi_S^* = \sum_{i \in S} \Pi_i^*(S)$ through transfers t_i such that

the “corrected” payoffs are $\Pi_i^*(S) + t_i$ with $\sum_{i \in S} t_i = 0$. If t_i is positive a coalition member receives transfers from other members and if it is negative a member makes a contribution to other members. The sum of transfers is zero, implying that there are no outside sources to pay for transfers.

In the case of transfers, many schemes are perceivable, which typically lead to different sets of stable coalitions. In order to avoid this sensitivity, we follow the concept of an almost ideal sharing scheme (AISS) proposed by Eyckmans and Finus (2009) with similar notions discussed for instance in Fuentes-Albero and Rubio (2009), McGinty (2007) and Weikard (2009). They basically argue that if and only if:

$$(4) \quad \text{Potential internal stability: } \sum_{i \in S} \Pi_i^*(S) \geq \sum_{i \in S} \Pi_i^*(S \setminus \{i\})$$

$$(5) \quad \text{Potential external stability: } \sum_{j \notin S} \Pi_j^*(S) > \sum_{j \notin S} \Pi_j^*(S \cup \{j\}) \text{ for all } j \notin S.$$

hold, then there exists a transfer system which makes S internally and externally stable. For instance, potential internal stability means that the total payoff of coalition S exceeds the sum of free-rider payoffs. In other words, the surplus $\sigma(S) = \sum_{i \in S} \Pi_i^*(S) - \sum_{i \in S} \Pi_i^*(S \setminus \{i\})$ is positive. Eyckmans and Finus (2009) show that such a transfer scheme is the almost ideal sharing scheme (AISS) which gives each coalition member his/her free-rider payoff $\Pi_i^*(S \setminus \{i\})$ plus a share of the surplus $\sigma(S)$. For our purposes, the following properties of AISS are important. First, among those coalitions which are potentially internally stable, the coalition with the highest aggregate welfare is stable, i.e. internally and externally stable. This explains the word “ideal”. Second, this coalition with the highest aggregate welfare may not be the grand coalition because free-rider incentives may be too strong (i.e. the surplus is negative). This explains the word “almost”. Finally, the aggregate welfare of the stable coalition which generates the highest aggregate welfare will never be lower with than without transfers, but most likely higher. That is, a smart transfer scheme like AISS will normally improve outcomes.

2.2 Three Types of Uncertainty

We now turn to the assumption about the uncertain parameters of the payoff functions, which are summarized in three *types of uncertainty*. Due to the complexity of coalition formation, the consideration of a particular payoff function, as well as the parameters that are uncertain and their distributions is required. In order to avoid the exclusive focus on the binary equilibrium choices “abate” or “not abate” in the second stage, as for instance in Kolstad (2007) and Kolstad and Ulph (2008), and to capture the first and second stage strategic effect from learning (for a definition see section 3.1), we consider a strictly concave payoff function (used for instance by Barrett 2006, Na and Shin 1998 and many others) which is still simple enough to derive analytical results:

$$(6) \quad \Pi_i = b_i \sum_{k=1}^n q_k - c_i \frac{q_i^2}{2}, \quad i \in N, \quad b_i > 0, \quad c_i > 0$$

where b_i is a benefit parameter, $b_i \sum_{k=1}^n q_k$ is the benefit from global abatement, c_i is a cost parameter, and $c_i \frac{q_i^2}{2}$ is the abatement cost from individual abatement.

Generally, the benefit as well as the cost parameters could be uncertain. In this chapter, we focus on benefit parameters and report briefly on results obtained in Finus and Pintassilgo (2012) for the cost parameters in section 3.3. Hence, we simplify the model, by dividing payoffs by the cost parameter c_i , define the benefit-cost ratio by $\gamma_i = b_i / c_i$, and hence the payoff function reads:

$$(7) \quad \Pi_i = \gamma_i \sum_{k=1}^n q_k - \frac{q_i^2}{2}, \quad i \in N, \quad \gamma_i > 0.$$

Henceforth, we call γ_i the benefit parameter. If this parameter is uncertain, then it is represented by the random variable Γ_i , with associated distribution f_{Γ_i} .

For all three types of uncertainty, uncertainty is symmetric. That is, all players know as much or little about their own as about their fellow players' payoff functions. We first lay out the specific assumptions

and then provide a wider interpretation. An overview is displayed in Table 1 and the formal description as well as generalizations are provided in Finus and Pintassilgo (2012 and 2013). All results and assumption are exactly those in Finus and Pintassilgo (2012).

Table 1 and 2 about here

Type 1: Uncertainty about the Level of Benefits

Uncertainty of type 1 is considered in Kolstad (2007) and Kolstad and Ulph (2008), which the authors call systematic uncertainty as it relates to a *common parameter*. All players have the same expectations ex-ante, and once uncertainty is resolved, all countries have the same benefit parameter ex-post, which we call symmetric realization. Important is that this type of uncertainty is de facto about the *level of the benefits* from global abatement.

Table 2 illustrates the implication with a simple example, which assumes only two players and a particular distribution of the benefit parameter, with takes two possible values with equal probability. Ex-post, all players have either a low value of $\gamma_i = 1$ or a high value $\gamma_i = 2$, but both have the same value and hence, ex-post, the sum of marginal benefits is either $\sum_{i=1}^2 \gamma_i = 2$ or $\sum_{i=1}^2 \gamma_i = 4$. The ex-ante expectation is $E(\Gamma_i) = 1.5$ and $\sum_{i=1}^2 E(\Gamma_i) = 3$. This can be viewed as an urn with tickets representing all possible values of the benefit parameter γ_i . One player is chosen randomly and draws a ticket; this value applies to all.

Type 2: Uncertainty about the Distribution of Benefits

Uncertainty of type 2 is considered in Na and Shin (1998) and relates to *individual parameters*. Though expectations about the benefit parameters are again symmetric, their realizations are asymmetric among players. More specifically, assume that there are n different tickets γ_i in an urn with values from 1, 2, ..., n ; each player selects a ticket without replacement. Hence, the realization of the benefit parameter is asymmetric. Table 2 illustrates this for our simple example. If player 1 draws $\gamma_1 = 1$, then player 2 will have

$\gamma_2 = 2$ and vice versa. This implies that the sum of marginal benefits is known ex-ante, $\sum_{i=1}^2 E(\Gamma_i) = \sum_{i=1}^2 \gamma_i = 3$ but not the individual marginal benefits, γ_i , with expected value $E(\Gamma_i) = 1.5$.

Different from Na and Shin (1998), we consider an arbitrary number of players and not only three players. We assume a discrete uniform distribution of the benefit parameters over all permutations of vector $(1, 2, \dots, n)$. Consequently, the sum of marginal benefits is fixed and consequently its variance is zero. Hence, uncertainty is purely about the *distribution of the benefits* from global abatement as the level of global benefits is known and constant $\sum_{i=1}^n E(\Gamma_i) = \sum_{i=1}^n \gamma_i = \frac{n^2 + n}{2}$. Note that the benefit vector can be viewed as different shares of the global benefits from abatement. Because the realization of the benefit parameters are asymmetric ex-post, it will turn out that transfers are useful and make a difference to the outcome when players learn this value (which is the case under our scenario of full learning, which is explained below).

Type 3: Uncertainty about the Level and Distribution of Benefits

Uncertainty of type 3 is a combination of the previous two types of uncertainty and hence there is uncertainty about *common and individual parameters*. This translates in our setting into uncertainty about the *level and distribution* of the benefits from global abatement. This is captured by assuming that, for each player, the benefit parameter follows a discrete uniform distribution over the values of vector $(1, 2, \dots, n)$, which, contrary to case 2, is independent from the benefit parameters of the other players. This can be viewed as players drawing tickets from an urn with n different tickets with values $1, 2, \dots, n$, not without but with replacement. Hence, the sum of marginal benefits is uncertain with positive variance which is larger than for uncertainty of type 2, but smaller than for uncertainty of type 1. Again, Table 2 is useful for illustrative purposes, which shows that symmetric but also asymmetric vectors of the benefit parameter may come about ex-post. On average, the asymmetry across players is larger than for uncertainty of type 1 but smaller than for uncertainty of type 2.

Interpretation of the Three Uncertainty Cases

All three types of uncertainty capture important aspects of the uncertainty surrounding climate change. There is much uncertainty about the absolute level of the benefits from reduced damages but also much debate about their regional distribution: which countries will be suffering more from climate change? Hence, uncertainty of type 3 represents the most comprehensive assumption, but type 1 and 2 are useful benchmarks in order to isolate effects.

2.3 Three Learning Scenarios

Following Kolstad and Ulph (2008), we assume risk-neutral agents as players are governments and not individuals, and distinguish three simple learning scenarios: 1) full learning, 2) partial learning and 3) no learning. The impact for coalition formation are illustrated in Table 3.

Table 3 about here

Full Learning (abbreviated FL) can be considered as a benchmark case in which players learn about the true parameter values before taking the membership decision in the first stage. Hence, uncertainty is fully resolved at the beginning of the game. Hence, this scenario could also be called certainty. For Partial Learning (abbreviated PL) it is assumed that players decide about membership under uncertainty but know that they will learn about the true parameter values before deciding upon abatement levels in the second stage. Hence, the membership decision is based on expected payoffs, under the assumption that players will take the correct decision in the second stage. We could also call this scenario partial uncertainty. Finally, under No Learning (abbreviated NL) also the abatement decision has to be taken under uncertainty. That is, players derive their abatement strategies by maximizing their expected payoffs. The membership decisions are also taken based on expected payoffs, though these expected payoffs differ from those under partial learning, given that less information is available under no learning in the second stage. We could call this scenario also uncertainty.

Full learning is certainly an optimistic and no learning a pessimistic benchmark about the role of learning in the context of climate change. Partial learning approximates (because beliefs are not updated in a Bayesian sense) the fact that information becomes available over time. For instance, between the signature of the Kyoto Protocol in 1997, its entry into force in 2005, and the target period of implementation 2008-12 (commitment period), more information has emerged, as documented by various updated issues of IPCC reports.

3. Results

3.1 Preliminaries

In this section, the main results of the coalition formation game are presented and discussed. Detailed derivations are found in Finus and Pintassilgo (2012). Note that for a sensible comparison across different scenarios, we measure aggregate welfare in terms of expected values, i.e. ex-ante. For FL, with no uncertainty, each possible realization is assumed to be equally likely and thus expected welfare is de facto an average welfare.

In order to understand the intuition of the subsequent results, it is useful to define what we call the first and second stage effect from learning. We say that the *first stage effect* from learning is positive (negative) if the size or the composition of the coalition leads to higher (lower) global welfare the more players learn. In our setting, a larger average coalition size usually results in higher global welfare. However, in some special cases, it is not the average size of stable coalitions which matters but the composition of coalition members. For instance, consider a coalition S of size m and an asymmetric realization of the benefit parameter γ_i . The highest global payoff is obtained if the m countries are those with the highest γ_i -values: the higher the γ_i -values of members, the higher their implemented abatement level and hence global abatement. Finally, we say that the *second stage effect* from learning is positive (negative) if for a given generic coalition S , global welfare is higher (lower) the more players learn, which is the result of the choice of equilibrium abatement levels.

The idea is to separate effects in both stages, though, for the overall result, effects in both stages matter. Three comments are in order. First, it is evident from Table 3 that partial and full learning are identical in the second stage. Second, due to backwards induction, the first stage effect cannot be completely isolated from what players do in the second stage. Nevertheless, thinking about first and second stage effects is useful when discussing the intuition of our results. Third, proofs use information about first and second stage effects. If a learning scenario performs (weakly) better with respect to both stage, it is straightforward to conclude that it performs (weakly) better overall. If effects in both stages go in opposite directions, the relative importance of effects needs to be weighted.

3.2 Main Results

We now turn to our main results, which are summarized in three Propositions according to the three types of uncertainty.

Proposition I: Uncertainty of Type 1 (uncertainty about the level of benefits)

For uncertainty of type 1, under the full, partial, and no learning scenario, equilibrium expected total payoffs are ranked as follows: $FL = PL > NL$

Thus, if there is only uncertainty about the level of the benefits from global abatement, “learning is good” in terms of aggregate welfare.ⁱⁱⁱ A detailed analysis would reveal that the first stage effect from learning is neutral, the size of stable coalitions for all three scenarios of learning is the same, namely $m^* = 3$. So what matters is the second stage effect from learning, which is positive for full and partial learning compared to no learning. Under no learning, abatement levels are chosen based on expected benefit parameter values. If the realized benefit parameters are all high (low), total expected abatement is too high (low) compared to the true parameter values. In other words, under no learning, there is systematic over- or undershooting of abatement, which is costly, and leads to lower aggregate welfare than under full and partial learning.

Proposition II: Uncertainty of Type 2 (uncertainty about the distribution of benefits)

For uncertainty of type 2, under the full, partial and no learning scenario, equilibrium expected total payoffs are ranked as follows:

$$\begin{array}{l}
 \text{No Transfers} \quad \left\{ \begin{array}{ll} NL = PL > FL & \text{if } n = 3 \\ NL > PL > FL & \text{if } n \geq 4 \end{array} \right. \\
 \\
 \text{Transfers:} \quad \left\{ \begin{array}{ll} FL = PL = NL & \text{if } n = 3 \\ NL > FL = PL & \text{if } 4 \leq n \leq 8 \\ NL > FL > PL & \text{if } n = 9 \\ FL > NL > PL & \text{if } n \geq 10 \end{array} \right.
 \end{array}$$

The main conclusion one can draw from Proposition II is that the more players learn, the lower will be global welfare if there are no transfers. This result can be somehow reversed with transfers. The intuition is the following.

First assume no transfers. Consider the first stage of coalition formation in which players choose their membership. For uncertainty of type 2, players are ex-ante symmetric though ex-post asymmetric. For partial and no learning this does not affect coalition formation compared to uncertainty of type 1 because players take their membership decisions based on expected payoffs which are symmetric (see table 3). Hence, $m^* = 3$. This does not apply to full learning. Members of a coalition receive different payoffs. As we assumed symmetric abatement cost function, cost-effectiveness requires that all coalition members contribute the same abatement level. However, members benefit to a different extent from this joint action as they have different benefit parameters γ_i . Those with high γ_i -values benefit more on average and those with low γ_i -values less on average. Thus, the low γ_i -value countries have an incentive to leave the coalition. This implies smaller stable coalitions than under symmetry. Hence, $m^* = 1$ if $n = 3$ and $m^* = 2$ if $n \geq 4$. In other words, the first stage effect from learning is negative for full learning but neutral for partial learning. Hence, regarding the first stage, assuming no transfers, the ranking is $NL = PL > FL$.

Consider now the second stage of coalition formation in which players chose their abatement levels. Consider the simplest case where all players are singletons and hence all behave non-cooperatively. For payoff function (6) the first order conditions, implying that individual marginal benefits are equal to marginal costs, delivers $q_i^* = \gamma_i$ under full and partial learning and $q_i^* = E[\Gamma_i]$ under no learning. Let us simplify things even further and consider only two countries and the realized parameter values as listed in Table 2. Then under full and partial learning either $q_1^* = 1$ and $q_2^* = 2$ or $q_1^* = 2$ and $q_2^* = 1$ whereas under no learning $q_1^* = 1.5$ and $q_2^* = 1.5$. The average or total abatement level is the same under all scenarios of learning but the individual abatement levels are asymmetric under full and partial learning but symmetric under no learning. For symmetric abatement cost function, cost-effectiveness requires symmetry, and hence total abatement costs under full and partial learning are higher (and hence total welfare lower) than under no learning. Hence, in a strategic setting, more information can lead to worse outcomes overall. Specifically, the second stage effect from learning is negative.

In our context, this simple example illustrating the second stage effect of learning generalizes in the following sense. First, the same result holds if we consider not only two but any number of countries. Second, the result holds not only if all players are singletons but for any coalition structure where there is a coalition S with m members and $n-m$ single players. If and only if S is the grand coalition, then the second stage effect from learning is neutral. Hence, regarding the second stage, the ranking is $NL \geq PL = FL$ with strict inequality if the coalition is not the grand coalition.

Thus, the overall result summarized in Proposition II, under the assumption of no transfers, with ranking $NL \geq PL > FL$ with strict inequality if $n \geq 4$, $PL > FL$ is (exclusively) due to the negative first stage effect from learning and $NL > PL$ is (exclusively) due to the negative second stage effect from learning. ($n=3$ is an exception because under PL and NL the grand coalition forms in which case the second stage effect from learning is neutral, which explains $NL = PL$ in this particular case.) Consequently,

we can conclude that the ranking $NL > FL$ is due to a negative effect from learning in the first and second stage.

The remaining question is: what do transfers change? They do not change the choice of equilibrium abatement levels and hence cannot change the second stage effect from learning. Hence, regarding the second stage, the ranking $NL \geq PL = FL$ still holds, with strict inequality if not the grand coalition forms. However, transfers change the first stage effect from learning provided this is due to asymmetry. Hence, transfers will make no difference to no and partial learning and consequently the ranking between these two scenarios will not change. We recall, regarding the first stage, this means $NL = PL$. This is different for full learning. First, transfers imply for full learning at least the same coalition size than for no and partial learning. Second, if asymmetries between players are pronounced enough, which increases with the number of players for the considered distribution (and hence results in Proposition II depend on n), even larger coalitions than under no and partial learning can be stable.^{iv} That is, the first stage effect from learning is neutral or even positive. According to Proposition II, the first stage effect from learning is strictly positive for $n \geq 9$, which explains the overall ranking $FL > PL$ if $n \geq 9$, and it is sufficiently strong for $n \geq 10$ to compensate the negative second stage effect from learning compared to no learning, which explains the overall ranking $FL > NL$ if $n \geq 10$. Thus, asymmetry is a burden without transfers but becomes an asset if a smart transfer scheme is used to balance asymmetries. The intuition is that the relative gains from cooperation accruing to coalition members compared to no cooperation increases with asymmetry, making it more attractive to join the coalition than remaining an outsider.

Taken together, we generalize the negative result of Na and Shin (1998) about the role of learning by considering more than three players and including the intermediate case of partial learning in the analysis. Even more important, we qualify their conclusion by considering transfers and showing that this conclusion can be reversed, at least for full learning.

Finally, like for uncertainty of type 2, for uncertainty of type 3, players are ex-ante symmetric but ex-post asymmetric. The average degree of asymmetry ex-post is positive, therefore larger than for

uncertainty of type 1, but smaller than for type 2. Not surprisingly, this improves upon the relative performance of full learning compared to uncertainty of type 2, but weakens it compared to type 1 if there are no transfers. With transfers, like in case 2, heterogeneity becomes an asset under full learning. The second stage effect from learning is positive like for uncertainty of type 1 which explains why partial learning always performance better than no learning and that this may even be true for full learning. Even without transfers, the first stage effect from learning under full learning can be positive compared to partial and no learning because not only the size but also the composition of coalition members matters.

Proposition III: Uncertainty of Type 3 (uncertainty about the level and distribution of benefits)

For uncertainty of type 3, under the full, partial, and no learning scenario, equilibrium expected total payoffs are ranked as follows:

$$\begin{array}{ll}
 \text{No Transfers} & \begin{cases} PL > NL > FL & \text{if } n < 29 \\ PL > FL > NL & \text{if } 29 \leq n < 32 \\ FL > PL > NL & \text{if } n \geq 32 \end{cases} \\
 \text{Transfers} & \begin{cases} FL = PL > NL & \text{if } n = 3 \vee n = 4 \\ FL > PL > NL & \text{if } n \geq 5 \end{cases}
 \end{array}$$

If we view uncertainty of type 3 as the most relevant case of actual negotiations because it captures uncertainty about the level and the distribution of the benefits from cooperation at the same time, both relevant in climate change, then our results come to a far less negative conclusion than the previous literature. Even in a strategic context, more information must not necessarily be detrimental to the self-enforcing provision of a public good. However, the larger the uncertainty about the distribution of the gains from cooperation, the more important it is to hedge against free-riding through an appropriate transfer scheme.

3.3 Extensions

So far the analysis focused on uncertainty about the benefits from abatement. A general conclusion was that the second stage effect from learning is only negative if there is pure uncertainty about the distribution of

benefits (uncertainty of type 2). Already if there is some uncertainty about the level of benefits (uncertainty of type 3) did this effect become positive and it was also positive if there is pure uncertainty about the level of benefits. If we were to assume the same three types of uncertainty about costs instead of benefits, it can be shown that the second stage of learning is always positive.

Another conclusion was that the first stage effect of learning can be negative for full learning compared to no and partial learning if the ex-post realization of the benefit parameters is sufficiently asymmetric under no transfers, but this was just reversed under transfers. For uncertainty about the cost parameters, a similar result holds. Asymmetric costs can lead to smaller coalitions under full learning if not accompanied by transfers. Transfers neutralize this negative effect, but, different from uncertainty about benefits, cannot turn it into a positive effect.

Overall, regardless of the type of uncertainty about costs, partial learning leads always to better outcomes than no learning and this is also true for full learning provided transfers are used to balance possible asymmetries.

4. Conclusion

The results in this paper challenge the conclusion that the “veil of uncertainty” is conducive to the success of international agreements on climate change, as suggested in the previous literature. We showed that learning is only bad in a strategic context in very specific situations: there is pure uncertainty about the distribution of the benefits from abatement. However, this is most unlikely in the climate change context. Moreover, should the problem arise, it can be mitigated, fixed or even turned into an asset through an appropriate transfer scheme. Such a transfer scheme could also be replicated through an appropriate allocation of permits under a carbon-trading scheme. Hence, our message is not that we deny the possibility that in a strategic context we find: what is good for single players can turn out to be bad at the aggregate. To the contrary, it is important to be aware of this possibility. However, if this possibility is anticipated, appropriate measures can be taken. Moreover, we derive a very optimistic message: diversity can be an

asset if managed well. This is an important message because conventional wisdom would suggest that the more diverse agents are, the more difficult it is to establish cooperation. This wisdom is true if no compensation measures are taken, but can be reversed if the gains from cooperation are shared such that participation is attractive to all. For addressing climate change effectively, this means that more efforts are needed to design policies not only cost-effectively, but to be aware of their distributional implications. This is particularly important because the potential gains of cost-effective climate policies are larger the more countries differ regarding costs and benefits.

5. Future Research

The most obvious extensions for future research include the following items. First, in our model, if players learn, there is perfect learning. A more natural assumption would be gradual learning through the update of beliefs. Our assumption about partial learning does not really capture this. Second, learning is exogenous in our model. Therefore, learning-by-doing, learning-by-research and governmental policies that influence the stock of knowledge could be part of model. Third, the first and second stage of coalition formation are one-shot decisions. In reality, countries can revise their membership decisions as well as their abatement decisions. Fourth, we assumed that all players hold the same information, whereas in reality there is dispute about scientific evidence with the IPCC for instance aiming at moderating this dispute. It would be interesting to analyze the effect of such institutions on the outcome of climate negotiations.

Acknowledgements

This chapter was written while M. Finus was visiting the Institute Henri Poincaré, Paris in January-February 2013. The hospitality of the Institute and financial support is gratefully acknowledged.

References

d'Aspremont, C., A. Jacquemin, J.J. Gabszewicz and J.A. Weymark (1983) On the Stability of Collusive Price Leadership, *Canadian Journal of Economics*, 16 (1), 17-25.

- Barrett, S. (2003) *Environment and Statecraft: The Strategy of Environmental Treaty-making*, Oxford University Press.
- Barrett, S. (2006) Climate Treaties and “Breakthrough” Technologies, *American Economic Review*, 96 (2), 22-25.
- Bloch, F. (2003), Non-cooperative Models of Coalition Formation in Games with Spillovers. In: C. Carraro C. (ed.), *The Endogenous Formation of Economic Coalitions*. Edward Elgar, Cheltenham, UK et al., ch. 2, pp. 35-79.
- Dellink, R. and M. Finus (2012), Uncertainty and Climate Treaties: Does Ignorance Pay? *Resource and Energy Economics*, 34, 565-584.
- Dellink, R., M. Finus and N. Olieman (2008), The Stability Likelihood of an International Climate Agreement. *Environmental and Resource Economics*, 39(4), 357-377.
- Eyckmans, J. and M. Finus (2009) An Almost Ideal Sharing Scheme for Coalition Games with Externalities, *Stirling Discussion Paper Series, 2009-10, University of Stirling*.
- Finus, M. (2001), *Game Theory and International Environmental Cooperation*. Edward Elgar, Cheltenham, UK et al.
- Finus, M. (2003) Stability and Design of International Environmental Agreements: The Case of Transboundary Pollution, In: H. Folmer and T. Tietenberg (eds.), *International Yearbook of Environmental and Resource Economics 2003/4*, Edward Elgar, Cheltenham, UK, pp. 82-158.
- Finus, M. (2008), Game Theoretic Research on the Design of International Environmental Agreements: Insights, Critical Remarks and Future Challenges. *International Review of Environmental and Resource Economics*, 2(1), 29-67.
- Finus, M. and P. Pintassilgo (2012) International Environmental Agreements under Uncertainty: Does the Veil of Uncertainty Help? *Oxford Economic Papers*, 64, 736-764.

- Finus, M. and P. Pintassilgo (2013) The Role of Uncertainty and Learning for the Success of International Environmental Agreements, *Journal of Public Economics*, 103, 29-43.
- Finus, M. and B. Rundshagen (2003), Endogenous Coalition Formation in Global Pollution Control: A Partition Function Approach. In: Carraro, C. (ed.), *The Endogenous Formation of Economic Coalitions*. Edward Elgar, Cheltenham, UK et al., ch. 6, pp. 199-243.
- Fuentes-Albero, C. and Rubio, S.J. (2009), Can the International Environmental Cooperation Be Bought? *European Journal of Operation Research*, 2002, 255-264.
- IPCC, (2007) *Climate Change 2007: Synthesis Report*.
- Kolstad, C.D. (2007) Systematic Uncertainty in Self-enforcing International Environmental Agreements, *Journal of Environmental Economics and Management*, 53 (1), 68-79.
- Kolstad, C.D. and A. Ulph (2008) Learning and International Environmental Agreements, *Climatic Change*, 89 (1-2), 125-141.
- Kolstad, C.D. and A. Ulph (2011), Uncertainty, Learning and Heterogeneity in International Environmental Agreements. *Environmental and Resource Economics*, 50, 389-403.
- McGinty, M. (2007), International Environmental Agreements among Asymmetric Nations. *Oxford Economic Papers*, 59, 45-62.
- Na, S.-L. and H.S. Shin (1998) International Environmental Agreements under Uncertainty, *Oxford Economic Papers*, 50 (2), 173-185.
- Stern, N. (2006) *Stern Review: The Economics of Climate Change*, Report prepared for the HM Treasury in the UK.
- Weikard, H.-P. (2009), Cartel Stability under Optimal Sharing Rule. "The Manchester School", 77, 575-93.
- Yi, S.-S. (1997) Stable Coalition Structures with Externalities, *Games and Economic Behavior*, 20 (2), 201-237.

Yi, S.-S. (2003), Endogenous formation of economic coalitions: a survey of the partition function approach.
In: Carraro, C. (ed.), *The Endogenous Formation of Economic Coalitions*, Edward Elgar, Cheltenham,
UK et al., ch. 3, pp. 80-127.

Table 1: Three Types of Uncertainty about the Benefit Parameters

Type of Uncertainty	Interpretation of Parameters	Ex-ante Expectations of Parameters	Ex-post Realizations of Parameters
1) pure uncertainty about the level of benefits	common	symmetric	symmetric
2) pure uncertainty about the distribution of benefits	individual	symmetric	asymmetric
3) simultaneous uncertainty about the level and the distribution of benefits	common and individual	symmetric	symmetric and asymmetric

Table 2: Ex-Post Realization: Example

Type of Uncertainty	Possible Ex-Post Realizations	Feature
1) pure uncertainty about the level of benefits	(1,1) , (2,2)	symmetric
2) pure uncertainty about the distribution of benefits	(1,2) , (2,1)	asymmetric
3) simultaneous uncertainty about the level and the distribution of benefits	(1,1) , (2,2) , (1,2) , (2,1)	symmetric or asymmetric

Table 3: Three Scenarios of Learning

Stage 1: Membership

NL	PL	FL
expected Π_i	expected Π_i	true Π_i

Stage 2: Abatement Decision

NL	PL	FL
expected γ_i	true γ_i	true γ_i

Notes

-
- i For possible other coalition formation games, see for instance Finus and Rundshagen (2003).
- ii Hence, one could also talk about a singleton coalition, i.e. a coalition with just one player. However, we find it easier to reserve the term coalition for a coalition of at least two players and refer to all players not belong to this coalition as singletons or non-coalition members.
- iii This result is in stark contrast to Kolstad (2007) and Kolstad and Ulph (2008). Since the technical details to explain this difference are quite involved, we refrain from discussing this issue here but refer the interested reader to Finus and Pintassilgo (2013).
- iv For the uniform distribution of the benefit parameter γ_i , the variance of the realized parameter values increases with the number of players n .